

WHAT IS CLAIMED IS:

1. A method of automatic information filtering for identifying inappropriate information among various information provided through Internet and blocking presentation of identified inappropriate information, comprising the steps of:

entering an HTML (HyperText Markup Language) information provided through the Internet;

10 judging whether a URL (Uniform Resource Locator) of said HTML information entered from the Internet is a top page URL or not, the top page URL being a URL ending with a prescribed character string defined according to a URL hierarchical structure by which each URL is constructed;

15 extracting words appearing in information indicated by the top page URL and carrying out an automatic filtering to judge whether said information indicated by the top page URL is inappropriate or not according to the words extracted from said information indicated by the top page URL, when said URL of said HTML information is the top page URL;

20 registering an upper level URL derived from the top page URL into an inappropriate upper level URL list and blocking presentation of said information indicated by the top page URL, when said information indicated by the top page URL is judged as inappropriate by the automatic filtering, the upper level URL being derived from the top page URL by keeping a character string constituting the top page URL only up to a rightmost slash;

25 comparing said URL of said HTML information with each URL registered in the inappropriate upper level URL list and judging whether there is any matching URL in the inappropriate upper level URL list when said URL of said HTML information is not the top page URL, and blocking presentation of information indicated by said URL of said

HTML information when there is a matching URL in the inappropriate upper level URL list, the matching URL being one upper level URL whose character string is contained in said URL of said HTML information;

5 extracting words appearing in said information indicated by said URL of said HTML information, and carrying out the automatic filtering to judge whether said information indicated by said URL of said HTML information is inappropriate or not according to the words extracted
10 from said information indicated by said URL of said HTML information, when there is no matching URL in the inappropriate upper level URL list; and

blocking presentation of said information indicated by said URL of said HTML information when said information
15 indicated by said URL of said HTML information is judged as inappropriate by the automatic filtering.

2. The method of claim 1, further comprising the steps of:

20 registering in advance URLs that provide inappropriate information in an inappropriate URL list; and

carrying out a third part rating based filtering for comparing said URL of said HTML information with each URL registered in the inappropriate URL list and judging
25 whether there is any matching URL in the inappropriate URL list, and blocking presentation of said information indicated by said URL of said HTML information when there is a matching URL in the inappropriate URL list.

30 3. An automatic information filtering apparatus for identifying inappropriate information among various information provided through Internet and blocking presentation of identified inappropriate information, comprising:

35 an input unit for entering an HTML (HyperText Markup

Language) information provided through the Internet;

5 a top page URL judging unit for judging whether a URL
(Uniform Resource Locator) of said HTML information entered
from the Internet is a top page URL or not, the top page
URL being a URL ending with a prescribed character string
defined according to a URL hierarchical structure by which
each URL is constructed;

10 a first automatic filtering unit for extracting words
appearing in information indicated by the top page URL and
carrying out an automatic filtering to judge whether said
information indicated by the top page URL is inappropriate
or not according to the words extracted from said
information indicated by the top page URL, when said URL of
said HTML information is the top page URL;

15 an inappropriate upper level URL list registration
unit for registering an upper level URL derived from the
top page URL into an inappropriate upper level URL list and
blocking presentation of said information indicated by the
top page URL, when said information indicated by the top
20 page URL is judged as inappropriate by the automatic
filtering, the upper level URL being derived from the top
page URL by keeping a character string constituting the top
page URL only up to a rightmost slash;

25 an inappropriate URL judging unit for comparing said
URL of said HTML information with each URL registered in
the inappropriate upper level URL list and judging whether
there is any matching URL in the inappropriate upper level
URL list when said URL of said HTML information is not the
top page URL, and blocking presentation of information
30 indicated by said URL of said HTML information when there
is a matching URL in the inappropriate upper level URL
list, the matching URL being one upper level URL whose
character string is contained in said URL of said HTML
information;

35 a second automatic filtering unit for extracting words

appearing in said information indicated by said URL of said HTML information, and carrying out the automatic filtering to judge whether said information indicated by said URL of said HTML information is inappropriate or not according to the words extracted from said information indicated by said URL of said HTML information, when there is no matching URL in the inappropriate upper level URL list; and

an information presentation blocking unit for blocking presentation of said information indicated by said URL of said HTML information when said information indicated by said URL of said HTML information is judged as inappropriate by the automatic filtering.

4. The apparatus of claim 3, further comprising:

an inappropriate URL list registration unit for registering in advance URLs that provide inappropriate information in an inappropriate URL list; and

a third party rating based filtering unit for carrying out a third part rating based filtering for comparing said URL of said HTML information with each URL registered in the inappropriate URL list and judging whether there is any matching URL in the inappropriate URL list, and blocking presentation of said information indicated by said URL of said HTML information when there is a matching URL in the inappropriate URL list.

5. A computer usable medium having computer readable program codes embodied therein for causing a computer to function as an automatic information filtering apparatus for identifying inappropriate information among various information provided through Internet and blocking presentation of identified inappropriate information, the computer readable program codes include:

a first computer readable program code for causing said computer to enter an HTML (HyperText Markup Language)

information provided through the Internet;

5 a second computer readable program code for causing said computer to judge whether a URL (Uniform Resource Locator) of said HTML information entered from the Internet is a top page URL or not, the top page URL being a URL ending with a prescribed character string defined according to the URL hierarchical structure by which each URL is constructed;

10 a third computer readable program code for causing said computer to extract words appearing in information indicated by the top page URL and carry out an automatic filtering to judge whether said information indicated by the top page URL is inappropriate or not according to the words extracted from said information indicated by the top
15 page URL, when said URL of said HTML information is the top page URL;

20 a fourth computer readable program code for causing said computer to register an upper level URL derived from the top page URL into an inappropriate upper level URL list and block presentation of said information indicated by the top page URL, when said information indicated by the top page URL is judged as inappropriate by the automatic filtering, the upper level URL being derived from the top page URL by keeping a character string constituting the top
25 page URL only up to a rightmost slash;

30 a fifth computer readable program code for causing said computer to compare said URL of said HTML information with each URL registered in the inappropriate upper level URL list and judge whether there is any matching URL in the inappropriate upper level URL list when said URL of said HTML information is not the top page URL, and block presentation of information indicated by said URL of said HTML information when there is a matching URL in the inappropriate upper level URL list, the matching URL being
35 one upper level URL whose character string is contained in

said URL of said HTML information;

a sixth computer readable program code for causing
said computer to extract words appearing in said
information indicated by said URL of said HTML information,
5 and carry out the automatic filtering to judge whether said
information indicated by said URL of said HTML information
is inappropriate or not according to the words extracted
from said information indicated by said URL of said HTML
information, when there is no matching URL in the
10 inappropriate upper level URL list; and

a seventh computer readable program code for causing
said computer to block presentation of said information
indicated by said URL of said HTML information when said
information indicated by said URL of said HTML information
15 is judged as inappropriate by the automatic filtering.

6. The computer usable medium of claim 5, wherein the
computer readable program codes further include:

an eighth computer readable program code for causing
20 said computer to register in advance URLs that provide
inappropriate information in an inappropriate URL list; and

a ninth computer readable program code for causing
said computer to carry out a third part rating based
filtering for comparing said URL of said HTML information
25 with each URL registered in the inappropriate URL list and
judging whether there is any matching URL in the
inappropriate URL list, and blocking presentation of said
information indicated by said URL of said HTML information
when there is a matching URL in the inappropriate URL list.

30

7. A method of automatic information filtering for
identifying inappropriate information among various
information provided through Internet and blocking
presentation of identified inappropriate information,
35 comprising the steps of:

obtaining word weights of words to be used in judging whether presentation of each information should be blocked or not according to words contained in each information, by an automatic learning using learning data containing
5 inappropriate information whose presentation should be blocked and appropriate information whose presentation should not be blocked;

storing and managing the word weights in correspondence to respective words in a form of a weighted
10 word list;

extracting words contained in information entered from the Internet; and

reading out the word weight for each word extracted from said information, from the weighted word list,
15 calculating a total sum of the word weights of the words extracted from said information, and judging whether or not presentation of said information should be blocked or not according to the total sum.

20 8. The method of claim 7, wherein the automatic learning is based on a linear discrimination function that can discriminate the inappropriate information and the appropriate information on a vector space.

25 9. The method of claim 7, further comprising the steps of:

registering in advance URLs that provide inappropriate information in an inappropriate URL list; and

carrying out a third part rating based filtering for
30 comparing said URL of said HTML information with each URL registered in the inappropriate URL list and judging whether there is any matching URL in the inappropriate URL list, and blocking presentation of said information indicated by said URL of said HTML information when there
35 is a matching URL in the inappropriate URL list.

10. An automatic information filtering apparatus for identifying inappropriate information among various information provided through Internet and blocking presentation of identified inappropriate information, comprising:

a word weight learning unit for obtaining word weights of words to be used in judging whether presentation of each information should be blocked or not according to words contained in each information, by an automatic learning using learning data containing inappropriate information whose presentation should be blocked and appropriate information whose presentation should not be blocked;

a weighted word list storing unit for storing and managing the word weights in correspondence to respective words in a form of a weighted word list;

a word extraction unit for extracting words contained in information entered from the Internet; and

a judging unit for reading out the word weight for each word extracted from said information, from the weighted word list, calculating a total sum of the word weights of the words extracted from said information, and judging whether or not presentation of said information should be blocked or not according to the total sum.

11. The apparatus of claim 10, wherein the automatic learning is based on a linear discrimination function that can discriminate the inappropriate information and the appropriate information on a vector space.

12. The apparatus of claim 10, further comprising:

an inappropriate URL list registration unit for registering in advance URLs that provide inappropriate information in an inappropriate URL list; and

a third party rating based filtering unit for carrying

out a third part rating based filtering for comparing said URL of said HTML information with each URL registered in the inappropriate URL list and judging whether there is any matching URL in the inappropriate URL list, and blocking presentation of said information indicated by said URL of said HTML information when there is a matching URL in the inappropriate URL list.

13. A computer usable medium having computer readable program codes embodied therein for causing a computer to function as an automatic information filtering apparatus for identifying inappropriate information among various information provided through Internet and blocking presentation of identified inappropriate information, the computer readable program codes include:

a first computer readable program code for causing said computer to obtain word weights of words to be used in judging whether presentation of each information should be blocked or not according to words contained in each information, by an automatic learning using learning data containing inappropriate information whose presentation should be blocked and appropriate information whose presentation should not be blocked;

a second computer readable program code for causing said computer to store and manage the word weights in correspondence to respective words in a form of a weighted word list;

a third computer readable program code for causing said computer to extract words contained in information entered from the Internet; and

a fourth computer readable program code for causing said computer to read out the word weight for each word extracted from said information, from the weighted word list, calculate a total sum of the word weights of the words extracted from said information, and judge whether or

not presentation of said information should be blocked or not according to the total sum.

14. The computer usable medium of claim 13, wherein the
5 automatic learning is based on a linear discrimination function that can discriminate the inappropriate information and the appropriate information on a vector space.

10 15. The computer usable medium of claim 13, wherein the computer readable program codes further include:

a fifth computer readable program code for causing said computer to register in advance URLs that provide inappropriate information in an inappropriate URL list; and

15 a sixth computer readable program code for causing said computer to carry out a third part rating based filtering for comparing said URL of said HTML information with each URL registered in the inappropriate URL list and judging whether there is any matching URL in the
20 inappropriate URL list, and blocking presentation of said information indicated by said URL of said HTML information when there is a matching URL in the inappropriate URL list.

25

30

35